

DDSP: Differentiable Digital Signal Processing

Year: 2020 | Citations: 427 | Authors: Jesse Engel, Lamtharn Hantrakul, Chenjie Gu, Adam Roberts

Abstract

Most generative models of audio directly generate samples in one of two domains: time or frequency. While sufficient to express any signal, these representations are inefficient, as they do not utilize existing knowledge of how sound is generated and perceived. A third approach (vocoders/synthesizers) successfully incorporates strong domain knowledge of signal processing and perception, but has been less actively researched due to limited expressivity and difficulty integrating with modern auto-differentiation-based machine learning methods. In this paper, we introduce the Differentiable Digital Signal Processing (DDSP) library, which enables direct integration of classic signal processing elements with deep learning methods. Focusing on audio synthesis, we achieve high-fidelity generation without the need for large autoregressive models or adversarial losses, demonstrating that DDSP enables utilizing strong inductive biases without losing the expressive power of neural networks. Further, we show that combining interpretable modules permits manipulation of each separate model component, with applications such as independent control of pitch and loudness, realistic extrapolation to pitches not seen during training, blind dereverberation of room acoustics, transfer of extracted room acoustics to new environments, and transformation of timbre between disparate sources. In short, DDSP enables an interpretable and modular approach to generative modeling, without sacrificing the benefits of deep learning. The library will be made available upon paper acceptance and we encourage further contributions from the community and domain experts.